

Lossless Source Coding for Multiple Access Networks¹

Qian Zhao Michelle Effros

Department of Electrical Engineering, California Institute of Technology

Mail Code 136-93, Pasadena, CA 91125, U.S.A.

email: {qianz, effros}@caltech.edu

Abstract — A multiple access source code (MASC) is a source code designed for the following network configuration: a pair of jointly distributed information sequences $\{X_i\}_{i=1}^{\infty}$ and $\{Y_i\}_{i=1}^{\infty}$ is drawn i.i.d. according to joint probability mass function (p.m.f.) $p(x, y)$; the encoder for each source operates without knowledge of the other source; the decoder receives the encoded bit streams of both sources. The rate region for MASCs with arbitrarily small but non-zero error probabilities was studied by Slepian and Wolf. In this paper, we consider the properties of optimal truly lossless MASCs and apply our findings to practical truly lossless and near lossless code design.

I. INTRODUCTION

Given finite-alphabet sources $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ with joint p.m.f. $p(x, y)$, a zero-error instantaneous MASC for (X, Y) comprises encoders, $\gamma_X : \mathcal{X} \rightarrow \{0, 1\}^*$ and $\gamma_Y : \mathcal{Y} \rightarrow \{0, 1\}^*$, and decoder, $\gamma^{-1} : \{0, 1\}^* \times \{0, 1\}^* \rightarrow \mathcal{X} \times \mathcal{Y}$. For any input sequences x_1, x_2, \dots and y_1, y_2, \dots , decoder γ^{-1} perfectly reconstructs (x_1, y_1) from $\gamma_X(x_1), \gamma_X(x_2), \dots$ and $\gamma_Y(y_1), \gamma_Y(y_2), \dots$ by reading only $\gamma_X(x_1)$ and $\gamma_Y(y_1)$. If the decoder losslessly decodes X before decoding Y , then the problem reduces to coding Y with side information X available only at the decoder.

A class of zero-error instantaneous side-information codes is introduced in [1, 2]. Here Y is encoded so that $y, y' \in \mathcal{A}_x = \{y \in \mathcal{Y} : p(x, y) > 0\}$ for some $x \in \mathcal{X}$ implies that $\gamma_Y(y)$ is not a prefix of $\gamma_Y(y')$. The decoder first losslessly decodes X and then uses the value of X to determine the set $\{\gamma_Y(y) : y \in \mathcal{A}_x\}$ from which to decode Y . Since these codewords are prefix-free, the description of $Y \in \mathcal{A}_x$ is uniquely decodable given X .

This paper characterizes optimal zero-error instantaneous side-information MASCs. Extensions to the general zero-error and near-lossless MASC problems and efficient design algorithms appear in [3, 4].

II. DEFINITIONS

Symbols $y_1, y_2 \in \mathcal{Y}$ can be combined under $p(x, y)$ if $p(x, y_1)p(x, y_2) = 0$ for all $x \in \mathcal{X}$. The collection $\mathcal{G} \subseteq \mathcal{Y}$ is called a 1-level group for $p(x, y)$ if each pair $y_i, y_j \in \mathcal{G}$ can be combined under $p(x, y)$; the tree representation $\mathcal{T}(\mathcal{G})$ is a single node representing all members of \mathcal{G} . For any $y \in \mathcal{Y}$, (y) is a special case of a 1-level group. A 2-level group $\mathcal{G} = (\mathcal{R} : \mathcal{C}(\mathcal{R}))$ for $p(x, y)$ comprises a root \mathcal{R} and its children $\mathcal{C}(\mathcal{R})$, where \mathcal{R} is a 1-level group, $\mathcal{C}(\mathcal{R})$ is a set of 1-level groups, and for each $\mathcal{G}' \in \mathcal{C}(\mathcal{R})$, each pair $y_1 \in \mathcal{R}$ and $y_2 \in \mathcal{G}'$ can be combined under $p(x, y)$. In the tree representation $\mathcal{T}(\mathcal{G})$ for \mathcal{G} , $\mathcal{T}(\mathcal{R})$ is the root of $\mathcal{T}(\mathcal{G})$ and the parent of all subtrees $\mathcal{T}(\mathcal{G}')$ for $\mathcal{G}' \in \mathcal{C}(\mathcal{R})$. These ideas generalize to M -level group

$\mathcal{G} = (\mathcal{R} : \mathcal{C}(\mathcal{R}))$ for $M > 2$, where \mathcal{R} is a 1-level group and $\mathcal{C}(\mathcal{R})$ is a set of groups of $M - 1$ or fewer levels.

A partition $\mathcal{P} = \{\mathcal{G}_1, \dots, \mathcal{G}_m\}$ on \mathcal{Y} for $p(x, y)$ is a complete and non-overlapping set of groups. The tree representation $\mathcal{T}(\mathcal{P})$ for \mathcal{P} is built by linking the roots of $\mathcal{T}(\mathcal{G}_i)$, $i = 1, \dots, m$, to an empty root node \mathbf{r} . We refer to a 1-level group \mathcal{G} at node $\mathcal{T}(\mathcal{G})$ in $\mathcal{T}(\mathcal{P})$ using the vector \mathbf{n} describing the path from \mathbf{r} to $\mathcal{T}(\mathcal{G})$. Node \mathbf{n} has $K(\mathbf{n})$ children, denoted by $\mathbf{n}1, \mathbf{n}2, \dots, \mathbf{n}K(\mathbf{n})$, and \mathbf{n} 's subtree probability $Q(\mathbf{n})$ is the sum of the probabilities of \mathbf{n} 's members and descendants.

A matched code γ_Y for partition \mathcal{P} is a binary code where for any $\mathcal{T}(\mathbf{n}) \in \mathcal{T}(\mathcal{P})$ and any $y_1, y_2 \in \mathbf{n}$, $y_3 \in \mathbf{n}k$, $y_4 \in \mathbf{n}k'$ with $1 \leq k < k' \leq K(\mathbf{n})$: (1) $\gamma_Y(y_1) = \gamma_Y(y_2)$; (2) $\gamma_Y(y_1)$ is a prefix of $\gamma_Y(y_3)$ and $\gamma_Y(y_4)$; (3) $\{\gamma_Y(y_3), \gamma_Y(y_4)\}$ is prefix-free.

III. MAIN RESULTS

Theorem 1 Code γ_Y is a zero-error instantaneous side-information code for $p(x, y)$ if and only if γ_Y is a matched code for some partition \mathcal{P} of \mathcal{Y} for $p(x, y)$.

Theorem 2 The optimal matched code for arbitrary partition \mathcal{P} of \mathcal{Y} for $p(x, y)$ has description lengths $l^*(\mathbf{r}) = 0$ and $l^*(\mathbf{n}k) = l^*(\mathbf{n}) + \log_2(\sum_{j=1}^{K(\mathbf{n})} Q(\mathbf{n}j)/Q(\mathbf{n}k))$ for all $\mathcal{T}(\mathbf{n}) \in \mathcal{T}(\mathcal{P})$ and $k \in \{1, \dots, K(\mathbf{n})\}$ if these lengths are all integers.

The matched Huffman code $\gamma_Y^{(H)}$ for partition \mathcal{P} describes the step from $\mathcal{T}(\mathbf{n})$ to $\mathcal{T}(\mathbf{n}k)$ using a Huffman code for alphabet $\{1, \dots, K(\mathbf{n})\}$ and p.m.f. $\{Q(\mathbf{n}k)/\sum_{j=1}^{K(\mathbf{n})} Q(\mathbf{n}j)\}_{k=1}^{K(\mathbf{n})}$; concatenating the descriptions of the steps between \mathbf{r} and $\mathcal{T}(\mathbf{n})$ in $\mathcal{T}(\mathcal{P})$ gives $\gamma_Y^{(H)}(y)$ for all $y \in \mathbf{n}$. Matched arithmetic codes are similarly constructed from arithmetic step descriptions.

Theorem 3 The matched Huffman code for partition \mathcal{P} minimizes the expected rate over all matched codes for \mathcal{P} .

Using these results, we can calculate the optimal rate for each partition on \mathcal{Y} and choose the partition with the lowest optimal rate. The optimal code on that partition is the optimal zero-error instantaneous code for \mathcal{Y} .

REFERENCES

- [1] A. Kh. Al Jabri and S. Al-Issa, "Zero-error codes for correlated information sources," In *Proceedings of Cryptography*, pp. 17–22, Cirencester, UK, 1997.
- [2] Y. Yan and T. Berger, "On instantaneous codes for zero-error coding of two correlated sources," In *Proceedings of the IEEE International Symposium on Information Theory*, pp. 344, Sorrento, Italy, June 2000.
- [3] Q. Zhao and M. Effros, "Optimal code design for lossless and near-lossless source coding in multiple access networks," In *Proceedings of the IEEE Data Compression Conference*, pp. 263–272, Snowbird, UT, March 2001.
- [4] Q. Zhao and M. Effros, "Lossless and near-lossless source coding for multiple access networks," In preparation for submission to the *IEEE Transactions on Information Theory*.

¹This material is based upon work supported by NSF Award No. CCR-9909026 and Caltech's Lee Center for Advanced Networks.